图书情报工作
第64卷 第6期 2020年3月

专利文献技术要素识别方法研究*
——以纳米肥料领域为例

■ 李晓曼 张学福 宋红燕 孙巍
中国农业科学院农业信息研究所 北京 100081

摘 要：[目的/意义]专利是技术情报最可靠的来源之一，专利分析可以实现专利信息的挖掘和利用，辅助专利技术创新。技术演化分析是指技术主题产生、发展、转移和变化乃至湮灭的过程，如何对专利技术信息进行深度揭示是当前专利技术演化分析的研究重点，其关键在于技术要素信息的揭示。[方法/过程]通过典型专利剖析、特征识别构建技术要素识别框架，提出面向特定领域的基于专利文献的技术要素识别方法。[结果/结论]将提出的技术要素识别方法应用于纳米肥料领域，通过典型专利剖析确定5种技术要素：材料、产品、方法、功效和用途，并基于SAO结构及领域词表完成技术要素的识别。该方法能为技术演化分析奠定基础，能够从多角度展示领域技术信息。

关键词：技术演化 技术要素 SAO结构 纳米肥料 技术网络
分类号：G255.53
DOI：10.13266/j.issn.0252-3116.2020.06.008

1 引言

技术创新是促进经济发展的重要力量和衡量综合国力的重要标志[1]。专利文献是技术最有效的载体之一，因此，它们被用作分析技术发展的重要工具[2]。专利分析经常被用于创新研究，例如识别技术发展趋势[3-4]、识别新兴技术[5-7]、识别技术需求[8]和寻找潜在合作者[9-10]等。计算机语言学和文本挖掘技术的发展使得专利分析能够分析专利文献的文本内容，显著扩展了专利分析的范围，受到了相当大的关注[11-12]。通过对专利文献的文本内容进行深度挖掘可以梳理技术的发展脉络，能够为技术人员探究、回溯技术起源和发展提供可靠的帮助，并对确定科技优先领域、合理配置科技资源具有重要的意义[13]。

技术演化分析是指技术主题产生、发展、突破创新、转移和变化乃至湮灭的过程，是专利技术挖掘的重要内容[14]。当前技术演化分析的相关文献[15-23]对技术的研究仅局限在技术主题上，对技术拆分重组的重视不足，使研究者难以在更具体、细化的层次上对技术演化的特征规律展开研究。在文本分析的范畴中，通过测量文档之间的关键词的共现关系来识别研究领域的技术主题是进行技术演化分析的基础方法[18-19]。Y. G. Kim等研究了基于关键词向量的专利聚类方法，用于新兴技术的预测[20]。随着语义分析技术不断发展，J. Yoon等[21]基于SAO结构描述专利相似性的能力，通过构建专利语义网络来识别技术发展趋势。汪雪锋等[22]利用SAO结构信息挖掘专利文本中的技术信息来构建技术路线图；Z. Hu等[23]构建了三层分析单元（SAO基础语义单元、技术主题、技术范畴），通过统计主题在聚类簇的分布来绘制技术演化图。

当代技术日新月异，不同领域逐步出现交叉融合，为了更清晰地刻画技术演化过程，需要对技术进行细粒度的分析。技术通常是以系统形式存在的[24-25]，技术要素是指在技术演化分析的过程中涉及的技术不同分析维度，是构成技术系统并维持其发展演化的最小单元[26]。从技术系统的角度理解技术演化，其内容包含技术系统内部技术各要素自身的变化或发展及各技术要素之间的相互作用和共同发展[27]。借鉴TIRZ[28]

* 本文系国家社会科学基金项目"基于SAO语义结构的专利技术演化与预测研究"（项目编号：16BTQ067）和中国农业科学院科技创新工程项目"科技情报分析与评估创新团队"（项目编目：CAAS-ASTIP-2016-AII）研究成果之一。
作者简介：李晓曼（ORCID：0000-0001-5718-0047），博士研究生；张学福（ORCID：0000-0002-9387-7527），研究员，博士，博士生导师，通讯作者，E-mail：zhangxuefu@caas.cn；宋红燕（ORCID：0000-0002-2611-6023），助理研究员，博士后；孙巍（ORCID：0000-0001-5949-9123），研究员，硕士生导师。
收稿日期：2019-05-23 修回日期：2019-09-22 本文起止页码：59-68 本文责任编辑：易飞

技术进化路线的思想,基于技术演化分析进行技术预测的基础是获取不同技术要素的技术演化路径,技术演化路径的基础是技术要素及其关联关系,形成技术预测→技术演化→技术要素的依存关系。因此,识别技术要素及其技术要素之间的关联关系是细粒度地揭示相关技术演化情况的关键步骤,也可为未来技术预测奠定坚实基础。

本文面向技术演化分析的需求,以专利文献为分析对象,通过典型专利剖析初步确定领域技术要素及其特征,然后设计特征识别方法来自动识别技术要素的特征,并构建基于特征的技术要素识别框架。最后,以纳米肥料领域为研究对象,应用技术要素识别框架完成了技术要素的识别,并对框架进行了补充和完善。

2 相关研究工作

工(技术)欲善其事,必先利其器(技术要素)[25]。为了解决技术细粒度的问题,相关研究者尝试对技术信息进行分类,界定了不同的分析维度。宏观层面的技术信息需要通过一定的知识提炼、归纳才能得到,无法从原始专利文本中直接获取。基于问题-解决方案类型的技术分析主要依据 TRIZ 理论,其核心内容之一是矛盾与创新原理[29],矛盾是指发明中待解决的各种问题,而解决这些矛盾冲突的创新原理即解决方案。将专利技术信息表示成技术问题信息和技术解决方案信息,从而进行下一步的分析。胡正银等[30]基于 5W1H 模型将"What"表示专利文献的技术问题(Problem)信息,"How"表示解决技术问题的技术方案(Solution)信息,"Why"表示专利的实现技术功能(Function)与达成的技术效果(Effect)信息。付芸等[31]以技术问题和技术方案为分析维度,遴选创新解决方案。H. B. Kim 等[32]探讨了如何利用技术问题和技术方案来实现技术重用。

微观层面的技术信息来源于专利元数据字段,可以从专利文本中直接获取。S. Choi 等[33]在构建技术树时设置了产品、技术和功能 3 个维度;S. Choi 等[11]在进行技术路线图绘制时对技术做了进一步的划分,设置了产品、技术、材料、技术属性和功能维度。翟东升等[34]通过对产品、功能、科学效应与功效等多维度的层级关系进行分析,构建技术树;X. Wang 等[35]在进行技术发展趋势识别分析时,划分了材料、技术、影响因素、组件、产品、目标和未来方向 7 个维度;汪雪锋等[22]构建了 6 层技术路线图(材料、技术、产品、目标、应用和影响因素);李倩[36]在进行专利新兴技术识别研究时,将短语类型、动宾组合及宾补组合的结构分为产品、方法、科学效应和技术属性;郭俊芳等[37]在进行技术形态识别研究时,将技术领域关键词分为技术属性、产品部件、部件材料、技术性能 4 个类别。

通过文献调研,可以发现,宏观层面上,技术的分析维度可以分为:技术问题和技术方案等。宏观层面的技术信息是经过后期分析总结得到的主题信息,单个的词或词组并不能表示其相关信息;微观层面上,技术的分析维度可以分为:技术、产品、方法、材料、组件/部件、功能、科学效应、技术属性、技术性能、应用领域、影响因素。微观层面的技术信息大多对应专利文本中的某个具体的技术短语或词组。其中,技术的主体信息包括 5 种技术要素,分别为:技术、产品、方法、材料、组件/部件;技术的效果信息包括 6 种技术要素,分别为:功能、科学效应、技术属性、技术性能、应用领域、影响因素。具体技术要素的解释如表 1 所示。

表 1 技术要素列表及含义信息

分析粒度	技术要素	解释
宏观层面	技术问题/研究问题	是指在当前技术领域存在的、技术发展过程中存在的问题[30-32]
	技术方案/研究方案	指在某一领域内,为了解决某一关键问题所提出的具体技术方案[30-32]
微观层面	技术	产品设计、制造和测试所需的技术,包括所有相关的过程、制造过程或方法[11, 22, 33, 35]
	产品	目标领域内的具体产品,是指可以从市场中获得并且技术活动所需的各种东西,包括有形的物品、无形的服务等[11, 22, 33-36]
	方法	方法是人们在技术实践中所利用的各种方法、程序、规则、技巧的总称。它帮助人们解决"做什么"、"怎么做"以及"怎样做得更好"的问题
	材料	某一技术方案所使用的关键材料[11, 22, 35]
	组件/部件	技术或产品的分支主题[35]
	功能	技术和产品所发挥的有利作用、效能[11, 34]
	科学效应	包含技术实现和产品功能实现过程中所用的科学原理、系统属性、现象与其工程应用等,其中效应指由某种动因或原因所产生的一种特定的科学现象[36]
	技术属性	测度技术发展成果的指标,通常与描述变化程度的动词连用[36]
	技术性能	衡量技术本身的特性[37]
	应用领域	指目标领域产品可以应用的具体领域[22]
	影响因素	技术领域发展面临的关键问题,该层次的问题可以被"技术层次"中的技术方案所解决[22]

总体来说,针对特定的技术领域,不同研究者根据分析的需要以及分析方法设置了不同的分析维度,对于技术要素的确定主要依靠分析者的经验选择,缺乏客观信息的支持。另外,对于技术要素的识别,仅提出了要素的特征标注方法如词性以及线索词等特征,并未形成规范化的流程设计方案,准确性和可操作性无法保证。

3 基于专利文献的技术要素识别

基于文献调研的技术要素为分析提供思路,可以作为技术要素识别的基础。技术要素的识别主要包括两部分内容:一是确定领域包含哪些技术要素,二是如何识别技术要素对应的关键词,并判断选定的技术要素是否可以覆盖领域关键词集,即满足每一种技术要素都有其对应的领域关键词以及每一个领域关键词都可划分为某一种技术要素。基于领域专利文献数据的技术要素识别方法主要分为 3 个步骤:

(1)典型专利剖析。选取特定研究领域的小样本量典型专利进行分析,剖析专利文本内容,挖掘领域所包含的技术要素。解析技术要素的特征信息如位置、语义等,标记技术要素的特征信息。

(2)特征识别。选取合适的方法识别特征信息,为自动化识别技术要素奠定基础。

(3)基于特征的技术要素识别框架。选取大样本领域数据进行实验,即判断每一种技术要素都有其对应的领域关键词以及每一个领域关键词都可划分为某一种技术要素。如果否,则有两种情况,一是出现了在抽样数据中未包含的技术要素,总结其特征信息,返回步骤(2);二是由于步骤(1)特征选取的不够导致技术要素未识别全,则需补充该技术要素的特征信息。该方法是一个动态的要素识别的过程,可以适应不同的学科领域。

基于领域专利文献数据的技术要素识别框架见图 1。

3.1 典型专利剖析

3.1.1 基于技术要素的专利知识表示

专利文献是专利知识的载体,专利技术方案是为了满足设计目的而采取的技术特征集合[38],可形式化为 PS = (SC, SR),其中 SC 为技术要素,SR 为技术要素的关联关系,如图 2 所示。从技术演化分析的目的出发,结合专利标题、摘要内容的具体表达结构进行特征选择,通过挖掘技术要素及其之间的关联关系,从而阐述发明的原理和发明的信息。因此,本节将专利剖析成技术要素信息,并对其特征进行总结。

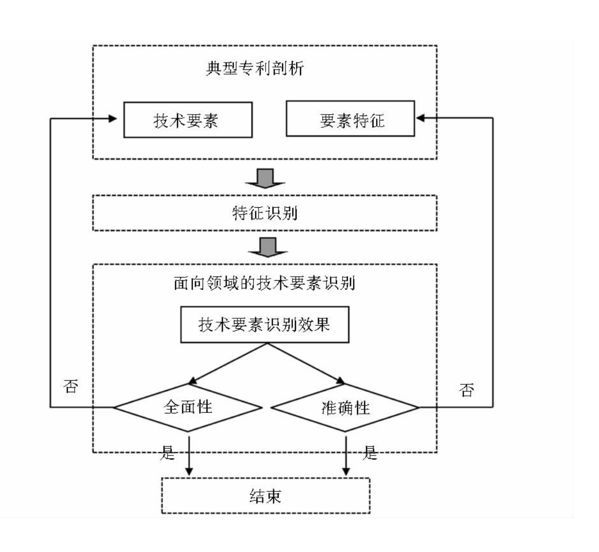


图 1 基于领域专利文献数据的技术要素识别框架

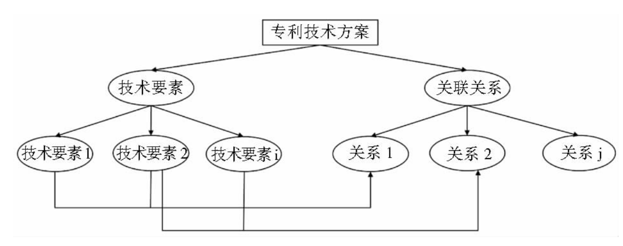


图 2 专利技术方案表示模型

3.1.2 领域专利剖析

以纳米肥料领域为例,在本节阐述领域专利剖析的内容。选取了纳米肥料领域 10 件专利作为典型专利进行剖析,提取领域关键词,总结关键词的含义,分为五大类。该领域涉及的技术要素有:产品、材料、方法、功效和应用领域。各技术要素的含义及其特征信息见表 2。

(1)对于产品技术要素,其位置信息没有明显特征,可出现在新颖性摘要 NO、用途摘要 USE 和创新性摘要 AD;对于其句法信息,经常出现在主语位置,但也可出现在宾语位置;其相关的动词也没有具体的特征,各种类型动词均有涉及;对于其语义信息,产品技术要素语义上是以“fertilizer”为中心词的短语。因此,利用语义信息可以识别产品技术要素相关的关键词。

(2)对于材料技术要素,其位置信息没有明显特征,可出现在 NO、USE、AD;对于其句法信息,经常出现在宾语位置;其相关的动词主要为包含类动词如 comprise 且处于 NO 位置,表示产品关键词的成分信息;对于其语义信息,材料技术要素语义上是以“material”“peel”“oxide”“acid”为中心词的短语。因此,利用位置信息、句法信息、动词信息和语义信息可以识别材料

技术要素相关的关键词。

（3）对于方法技术要素，其位置信息没有明显特征，可出现在 NO、USE、AD；对于其句法信息，经常出现在主语位置，但也可出现在宾语位置；其相关的动词也没有具体的特征，各种类型动词均有涉及，如作为方法关键词与包含类动词的宾语常为步骤类短语；对于其语义信息，方法技术要素语义上是以"method"为中心词的短语或为动名词与材料相结合的步骤类短语。因此，利用位置信息、句法信息、动词信息和语义信息可以识别方法技术要素相关的关键词。

（4）对于功效技术要素，其位置信息仅出现在 NO、AD；对于其句法信息，仅出现在宾语位置；对于其相关的动词，作为 improve、decrease 等描述变化动词的宾语，或作为 comprise 等包含类动词的宾语处于 AD 位置；对于其语义信息，语义上是以"cost""efficiency""safe""taste""speed"为中心词的短语。因此，利用位置信息、句法信息和动词信息可以识别功效技术要素相关的关键词。

（5）对于应用领域技术要素，其位置信息仅出现在 USE；对于其句法信息，仅出现在宾语位置；对于其相关的动词，作为 be useful for、be suitable for、used for 等动词的宾语；对于其语义信息，语义上是以"planting"为修饰语且与作物种类搭配的短语。因此，利用位置信息、句法信息和动词信息可以识别应用领域技术要素相关的关键词。

表 2　典型专利技术要素的定义及特征信息

技术要素	定义	特征信息	示例词
产品	在目标领域内的具体产品，即纳米肥料领域的具体产品	出现在 NO、USE、AD；经常出现在主语位置；各种类型的动词均有涉及；语义上是以"fertilizer"为中心词的短语	bio organic fertilizer、nano modified calcium fertilizer
材料	是指利用某种技术制作产品时所使用的关键材料，即制作纳米肥料相关产品所使用的关键材料	出现在 NO 、USE、AD；经常出现在宾语位置；常作为 comprise 等包含类动词的宾语；语义上是以"material""peel""oxide""acid"为中心词的短语	pig manure、microbial residue、trace element
方法	方法是人们在技术实践中所利用的各种方法、程序、规则、技巧的总称。包括在纳米肥料制作中的关键步骤以及各种制备方法	出现在 NO、USE、AD；主语和宾语位置均会出现；各种类型的动词均有涉及；语义上是以"method"为中心词的短语或为动词和材料组合的形式	preparing compound fertilizer、mixing strontium powder
功效	是指产品或材料的具体功能，即纳米肥料领域中促进植物生长、防虫、防病、提高作物产量和质量等	出现在 NO、AD；经常出现在宾语位置；作为 improve、decrease 等描述变化动词的宾语；语义上以"cost""efficiency""safe""taste""speed"为中心词的短语	efficiency、heavy metal、growth、good stability
应用领域	是指目标领域产品具有的具体用途及应用领域，纳米肥料领域的应用领域包括施用植物种类和施用位置两类	出现在 USE；经常出现在宾语位置；作为 be useful for、be suitable for、used for 等动词的宾语；语义上是以"planting"为修饰语且与作物种类搭配的短语	leafy vegetable、corn、underground tuber crop

3.2　特征识别

3.2.1　技术要素特征信息

将非结构化的以文本形式存在的专利信息转化为由一系列文本特征表示的数据集合，用以描述和替代自由文本并指导自动化识别要素信息。对于典型专利剖析确定的技术要素，通过分析各要素包含的特征信息，将技术要素的特征总结为以下 4 个，具体为：

特征 1：位置特征。对于德温特数据库中的专利信息，摘要是经过专家改写的结构化文本，不同部分的摘要表征不同的含义。如：应用领域常出现在用途摘要中，功效常出现在创新性摘要中。

特征 2：句法特征。句子的主要结构就是主谓宾结构，不同技术要素在句子中充当的含义不同。如：产品、材料和方法在主语和宾语位置都有可能出现，应用领域和功效仅会出现在宾语位置。

特征 3：动词特征。动词是支配其他词性成分的中心成分，所有受支配的成分都以某种依存关系从属于动词词性成分[39]。不同技术要素之间的关联关系可以通过动词的含义来展现，如：产品和材料之间的组成关系，对应的动词为"include"等。产品和应用领域之间的用途关系，对应的动词为"used for"等。

特征 4：语义特征。虽然名词的语义信息具有多义性，但在特定技术领域中，其语义相对稳定，减少了语言理解的歧义。因此，各技术要素语义的标记主要基于专利领域术语词典。如：出现"fertilizer"为中心名词的名词短语为产品。出现"material""peel""oxide""acid"等为中心名词的名词短语为材料。

3.2.2　SAO 结构与技术要素关联分析

从发明问题解决理论（TRIZ）的观点来看，SAO 三元组是一种描述功能实现和技术构成的模型。SAO 结构通过动词短语定义了两个技术关键词之间的关联关系，表示的是"技术关键词 – 关联关系 – 技术关键词"三元组。技术关键词之间的关联关系有多种，如：生物医学领域的基因与基因关系、蛋白质与蛋白质相互作

李晓曼，张学福，宋红燕，等. 专利文献技术要素识别方法研究——以纳米肥料领域为例[J]. 图书情报工作,2020,64(6):59－68.

用关系、基因与疾病的关系、基因与药物的关系、疾病与治疗的关系等[40]。微观层面的技术要素是技术关键词更抽象化的表示形式，通过对 SAO 结构中的技术关键词进行分析，识别技术关键词对应的技术要素。基于 SAO 结构的技术要素表示模型如图 3 所示：

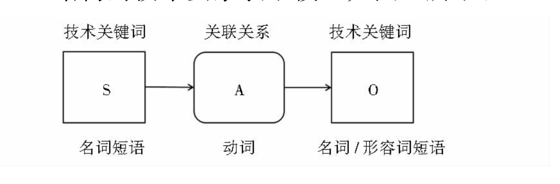


图 3　基于 SAO 结构的技术要素表示模型

3.2.3　基于 SAO 语义结构的特征识别

针对上述的特征信息，选取合适的方法标记特征信息，为构建自动化的抽取方法挖掘关键词与技术要素的对应关系奠定基础。

（1）对于特征 1，可通过专利文本的标识信息识别，在此不做详细阐述。

（2）对于特征 2，SAO 结构是由主谓宾结构构建的三元组，可以直观展示出某技术关键词是在句子中的成分。因此，利用关键词在 SAO 结构中的位置（即属于 S 还是 O）来判断其在句子中的成分。

（3）对于特征 3，对 SAO 结构的动词（A 部分）进行分类，构建专利动词语义框架来辅助 SAO 结构分类。专利动词知识库的建立是基于一定数量的专利文献训练集，通过对动词进行词频的统计，利用动词含义对常见的动词进行分类。该过程可以手工完成，也可以通过机器学习辅助完成。针对专利语义特点，利用相关动词的含义将 SAO 结构划分为：包含类（partitive）、功效类（function）、用途类（purpose）和相互作用类（interaction）。

包含类是表示包含关系的动词，所描述的是主语和宾语之间的包含层次关系，如产品和材料的包含关系，具体的动词有“contain”“consist”“have”等。功效类是表示属性变化的动词，所描述的是主语如何影响宾语的，这种类型的 AO 结构多被构造成诸如“improve efficiency”等形式，即提升类的动词和表示效率、产量等词的集合。用途类是表示使用关系的动词，所描述的是宾语是主语的应用领域，核心动词为“be used for”等形式。相互作用类主要是指主语和宾语之间的相互作用，如某种材料影响某种产品，某种方法可以用来生产某种产品。本文定义了各类型的核心动词参考表如表 3 所示：

表 3　关系类型和核心动词

关系类型	核心动词	含义	示例信息	
			技术要素关键词 1	技术要素关键词 2
包含类	comprise, have, include, contain, involve	表示主语和宾语之间的包含层次关系（S1 include S2）	产品（Water soluble fertilizer）	材料（modified soybean dreg）
功效类	improve, increase, reduce, promote degrade, achieve, realize, recycle, accelerate, alleviate, absorb, be easy to, be good for, stabilize	表示主语是如何改变宾语的技术属性或主语对宾语的影响（S1 is an objective for S2）	产品（Water soluble fertilizer）	功效（improve growth）
用途类	be used for, (be) used in	表示宾语是主语的应用领域（S1 is used in S2）	产品（Water soluble fertilizer）	应用领域（leafy vegetable）
相互作用类	prepare, provide, produce, obtain, form, develop, utilize	表示主语和宾语之间的相互作用（S1 has Interaction with S2）	材料（composite microbial inoculum）	材料（Aspergillus niger）

（4）对于特征 4，主要通过线索词补充领域词表，并利用 WordNet 语义词典辅助同义词合并。虽然名词具有多义性，但在特定技术领域中，其语义相对稳定，减少了语言理解的歧义。因此，技术要素的语义标记主要是基于专利领域的术语词表。识别规则通过建立名词短语的词性标记规律和外在边界词性标记规律来识别。如在纳米肥料领域，出现以“fertilizer”为中心名词的名词短语，其技术要素应标记为产品；出现以“material”“peel”“oxide”“acid”等为中心名词的名词短语，其技术要素应标记为材料。若术语词典中包含的词不全，则可进一步利用语义词典 WordNet 查询其语义信息，标记特征词和技术要素的对应关系。

3.3　技术要素识别框架

技术要素在专利文本中的位置、在句子中充当的成分、对应的动词类型以及语义方面不同的特征，可作为其判别的依据。特征分析阶段主要包括基于标题和摘要的语义特征信息标记，以专利标题和摘要的文本信息作为数据源进行语义标记的过程主要是实现从专利的语义信息中挖掘技术要素的特征。基于领域数据进行技术要素识别前，必须对专利文献进行预处理，包括：检索、下载和格式转换等，将专利文献中的各部分分段存储，形成具有统一数据格式的领域专利数据集。

基于 SAO 语义结构的技术要素识别方法通过 SAO 结构集来识别技术要素的特征信息。通过位置特

征、句法特征和动词特征可以实现功效和应用领域技术要素的识别。对于产品、材料和方法技术要素,需要进一步通过语义特征来识别。在此过程中,由于从典型专利剖析所利用的小数据量的专利文献集合到特定领域大数据量的专利文献集合,会存在部分技术词没有对应的技术要素。出现此问题的原因有两种:①由于专利动词知识库和领域词典中线索词不全导致部分技术词和技术要素未产生关联,需要人工总结其特征并补充相关知识库的信息;②由于出现了在抽样数据中未包含的技术要素,需要总结其特征信息,完善技术要素识别框架。基于特征的技术要素识别框架如图 4 所示:

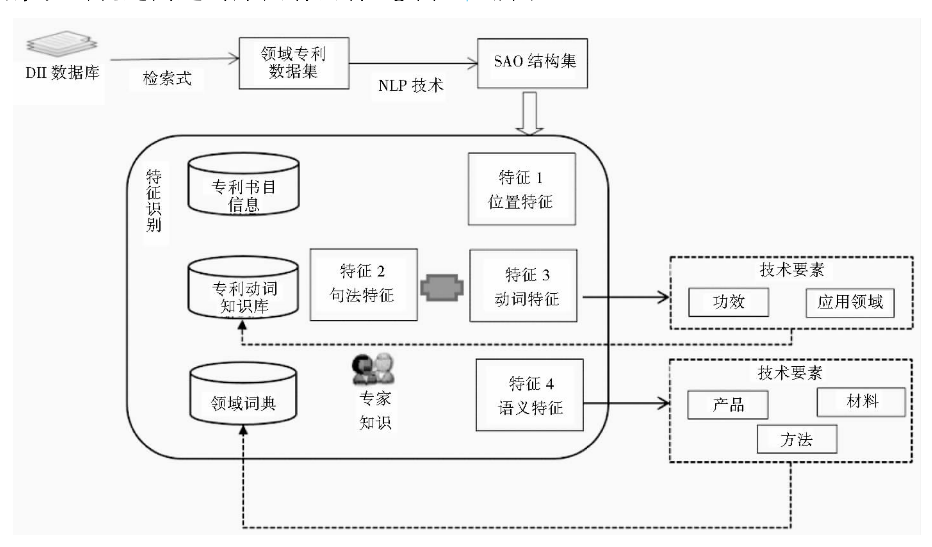


图 4 基于特征的技术要素识别框架

4 案例研究

本文以纳米肥料领域的专利数据作为数据源,展示技术要素的识别结果。纳米肥料是纳米技术在农业上的一个里程碑式的应用,它是用纳米材料技术构建、用医药微胶囊技术和化工微乳化技术改性及化学聚合而形成的全新肥料,包括纳米结构肥料与纳米材料包膜或胶结缓、控释肥料[41]。本文以纳米肥料领域相关专利为研究对象,检索日期为 2018 年 11 月 27 日。基于 DII 专利数据库检索获取了 779 件专利,以 2018 年的 46 件专利作为分析对象,完成案例分析部分。

4.1 纳米肥料领域典型专利剖析

选取了纳米肥料领域 10 件专利作为典型专利进行剖析,提取领域关键词,人工总结关键词的含义,分为五大类。该领域涉及的技术要素有:产品、材料、方法、功效和应用领域。对技术要素对应的领域关键词的相关特征进行了人工标注,标注的特征包括:技术要素的位置特征、句法特征、动词特征及语义特征。具体剖析的形式如表 4 所示:

表 4 "纳米肥料"专利技术要素特征剖析示例

专利号 CN108530210-A
内容 TI:Water-soluble fertilizer comprises modified soybean dregs, egg shell, plant ash, banana peel, apple peel, pear peel, nano cerium oxide, lutetium oxide, nano selenium, 5-aminolevulinic acid and bentonite.
USE - The water-soluble fertilizer is useful for planting corn

技术要素	位置特征	句法特征	动词特征	语义特征
产品(Water-soluble fertilizer)	TI、USE	主语	is useful for、comprises	名词性短语,线索词为 fertilizer
材料(modified soybean dregs…)	TI	宾语	comprises	名词性短语,线索词为 dregs、shell、oxide 等
应用领域(planting corn)	USE	宾语	is useful for	线索词为 planting

4.2 基于 SAO 语义结构的特征识别

4.2.1 SAO 结构抽取

文本首先利用自然语言处理技术(NLP)提取专利文本中的 SAO 结构。利用 CluseIE 借助 Java 开发环境抽取标题和摘要部分的 SAO 结构,利用 DDA 软件对原始 SAO 结构进行清洗,去除与领域无关的结构,得到待分析的 SAO 结构 857 个。

4.2.2 SAO 结构位置标记

对于从文本中提取出的 SAO 结构,将对应的位置信息以及专利号存储在 excel 中,以进行下一步的分

析。具体信息如表 5 所示：

表 5　2018 年纳米肥料专利文本抽取的 SAO 结构示例(部分)

专利号	S	A	O	位置
CN108314556 A	Selenium rich agricultural fertilizer	comprises	carboxylase	TI
CN108314556 A	Selenium rich agricultural fertilizer	comprises	wood vinegar	TI
CN108314556 A	Selenium rich agricultural fertilizer	comprises	mixed bacterial liquid	TI
CN108314556 A	Selenium rich agricultural fertilizer	comprises	nano selenium powder	TI
CN108314556 A	Selenium rich agricultural fertilizer	comprises	soy protein powder	TI
CN108314556 A	Selenium rich agricultural fertilizer	comprises	Bacillus megaterium	TI
CN108358700 A	Selenium enriched pitaya organic fertilizer	includes	nanoselenium plant nutrient agent	TI
CN108358700 A	Selenium enriched pitaya organic fertilizer	includes	straw ash	TI
CN108358700 A	Selenium enriched pitaya organic fertilizer	includes	urea	TI
CN108358700 A	Selenium enriched pitaya organic fertilizer	includes	animal feces	TI

4.2.3　基于动词语义框架的 SAO 结构分类

将 SAO 结构中的核心动词与构建的动词知识库中的动词进行匹配,完成了 676 个 SAO 结构的分类。SAO 结构语料中的核心动词 TOP10 部分信息如表 6 所示。对于剩余的 SAO 结构,利用专家知识将核心动词划分到具体的类别中,并将动词信息补充到动词知识库中。通过此部分研究,SAO 结构可分为 564 个包含类、197 个功效类、36 个用途类和 60 个相互作用类。

表 6　2018 年纳米肥料专利文本核心动词 TOP10(部分)

动词	词频	所属类别词
comprise	402	包含类
have	47	包含类
improve	46	功效类
used for	39	用途类
include	38	包含类
contain	26	包含类
involve	25	包含类
is	25	功效类
is useful for	21	用途类
prepared by	15	相互作用类

4.3　面向纳米肥料领域的技术要素识别框架

对于标题和新颖性部分的 SAO 结构,包含类 SAO 结构对应的 S 或 O 是产品、材料和方法技术要素,表示单一维度或多维度的技术要素层次关系;对于创新性摘要部分的 SAO 结构,包含类 SAO 结构对应的 S 部分是产品或方法技术要素,O 部分为功效技术要素;对于功效类 SAO 结构,其对应的 S 短语类型是产品、材料和方法技术要素,O 短语类型为功效技术要素;对于用途类 SAO 结构,其对应的 S 短语类型是产品、材料和方法技术要素,O 短语类型为应用领域技术要素;对于相互作用类 SAO 结构,表示不同技术要素之间的关联关系,如方法与产品之间的相互作用、方法与材料之间的相关作用等。因此,对于产品、材料和方法技术要素的识别,仅仅通过 SAO 结构的位置和类型信息无法全部确定,需要判断是否包含产品、材料和方法技术要素对应的线索词来建立与技术要素的对应关系。图 5 表示了纳米肥料领域技术要素识别框架。表 7 展示了各技术要素对应的特征值信息。

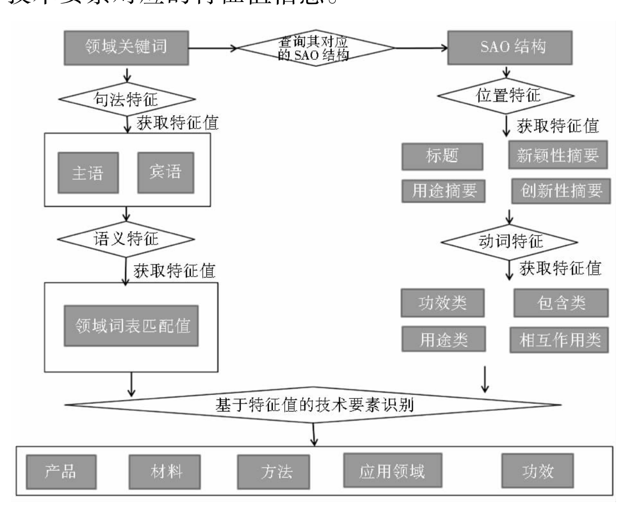


图 5　纳米肥料领域技术要素识别框架

表 7　基于位置和分类信息的技术要素特征值信息

关系类型	标题	新颖性摘要	用途摘要	创新性摘要
包含类	S:PMM O:PMM	S:PMM O:PMM	无	S:PMM AO/O:功效
功效类	S:PMM AO/O:功效	无	无	S:PMM AO/O:功效
用途类	S:PMM O:用途	无	S:PMM O:用途	无
相互作用类	S:PMM O:PMM	S:PMM O:PMM	S:PMM O:PMM	S:PMM O:PMM

注:PMM 指的是产品、材料和方法技术要素

4.4　技术要素识别结果

4.4.1　评价方法

本研究采用准确率和召回率作为指标对实验结果

进行评价,公式如下:

$$\text{Precision} = \frac{A \cap H}{A}$$

$$\text{Recall} = \frac{A \cap H}{H}$$

其中,Precision 表示准确率,Recall 表示召回率,A 表示实验方法识别出的结果个数,H 表示基准数据集合中的结果个数。本研究借助具有专业背景的人员对技术关键词对应的技术要素进行标注,作为基准数据集合。

4.4.2 识别结果分析

基于技术要素识别框架,首先通过 SAO 结构位置和分类信息共识别出 172 个功效关键词和 15 个用途关联词。通过纳米肥料领域词表以及线索词,确定其余的关键词对应的技术要素。通过此步识别,得到 50 个产品关键词、341 个材料关键词和 26 个方法关键词。通过专家知识进一步对纳米肥料领域涉及的技术要素进行适用性分析,验证了识别效果的准确性。首先,对于技术要素的划分,认为纳米肥料领域划分为 5 种要素是合理的,但是由于纳米肥料的主要技术是针对各材料含量的不同而实现的,后期可否增加材料含量相关的信息;对于技术关键词和技术要素的对应关系,目前的识别结果是合理的。SAO 结构的位置、分类信息统计情况如表 8 所示,技术关键词与技术要素的对应关系如表 9 所示:

表 8 SAO 结构位置和分类信息统计

关系类型	标题	新颖性摘要	用途摘要	创新性摘要
包含类	289	225	0	50
功效类	15	0	7	182
用途类	12	0	17	0
相互作用类	30	15	8	7

表 9 技术关键词与技术要素的对应关系

技术关键词	技术要素
organicic fertilizer, bio organicic fertilizer, complex fertilizer, environmental friendly fertilizer, foliar fertilizer, nitro fertilizer, selenium germanium enriched element fertilizer, silicon fertilizer, synergbetic fertilizer, water soluble fertilizer, …	产品
urea, plant ash, chitin, nano carbon, additive, bentonite, composite, copper sulfate, ferrous sulfate, humic acid, …	材料
method, produce organicic fertilizer, mix pig manure, prepare microelement nanochelate foliar fertilizer composite, achieve modified gas pressure, add calcium fertilizer, add magnesium fertilizer, add nano actingivated carbon, add phosphoricic acid, add potassium nitrate, …	方法
yield, quality, low cost, utilization rate, environmental friendly, synergbetic effect, growth, soil, efficiency, rich nutrient, safe, …	功效
aquaculture, Camellia oleifera, corn, cucumber, deodorizing fecal waste, fermentation fecal waste, flower, fruit tree pestss, grape, insect resbetance, …	应用领域

4.4.3 识别效果评估

技术要素识别结果评估如表 10 所示,经对比发现产品技术要素和材料技术要素识别效果较好,应用领域技术要素和功效技术要素次之,方法技术要素较差。一方面原因在于产品和材料技术要素多为名词或名词短语,线索词较为明显,因此识别效果较好;另一方面,方法技术要素对应的关键词多为名词或名词短语以及动名词与名词短语的组合。方法技术要素特征词集有待进一步完善,由于遴选具有显著方法类型特征的短语较为困难,对于方法技术要素的识别效果尚不完善。

表 10 技术要素识别结果评估

基准数据	人工标注结果	
技术要素	准确率	召回率
产品	86%	95.35%
材料	91.79%	98.43%
方法	100%	53.06%
应用领域	77.78%	93.33%
功效	91.86%	96.93%

5 结论

为满足细粒度技术演化分析的需求,本文以专利文献为分析对象,通过典型专利剖析初步确定领域技术要素及其特征,然后从位置、句法、动词和语义特征入手,设计了特征识别方法来自动识别技术要素的特征。基于此,本文构建了基于特征的技术要素识别框架,以纳米肥料领域为研究对象,应用技术要素识别框架完成了技术要素的识别,并对识别结果进行分析与评价,展示了技术要素识别方法的可行性和有效性。

本文提出的技术要素识别方法可以根据技术领域的特征构建不同的技术分析维度,基于特征识别技术要素以及技术要素之间的关联关系,可以从多角度、多层面展示专利的技术信息,方法具有灵活性和系统性。然而,本文仍存在以下不足之处:

(1)由于篇幅有限,本文仅对纳米肥料领域 2018 年的专利进行了深入分析,后期可以研究纳米技术在农业领域的应用中涉及到的所有产品领域,例如纳米疫苗、纳米兽药等。

(2)基于 SAO 结构完全自动化地识别技术要素还不够精确,各技术要素线索词遴选需要借助大量的领域数据。后续研究需要提升技术要素识别的自动化程度及准确性,进一步减少对专家知识的依赖性。

技术发展的日新月异,交叉融合是必然趋势。由此,细粒度分析已逐渐成为技术演化领域的研究重点。

识别个体技术要素当前所处研究阶段及其在不同阶段间转移扩散的时间线索和技术细节差异,能够有效支撑开展个体技术成熟度以及转移过程差异性分析等细粒度技术情报分析工作,为技术演化分析奠定数据基础。另外,在方法的应用上,后续研究需要探索如何真正利用技术要素构建技术演化所需的技术网络,如何将技术演化路径与 TRIZ 技术进化路径结合进行技术预测,达到发现潜在研究机会、辅助技术研发选题等情报工作目标。

参考文献：

[1] 陈亮,张志强. 技术演化研究方法进展分析[J]. 图书情报工作, 2012, 56(17): 59-66.

[2] KERR C I, MORTARA L, PHAAL R, et al. A conceptual model for technology intelligence[J]. International journal of technology intelligence and planning, 2006, 2(1): 73-93.

[3] PARK Y, YOON B, LEE S. The idiosyncrasy and dynamism of technological innovation across industries: patent citation analysis[J]. Technology in society, 2005, 27(4): 471-485.

[4] ZHANG L. Identifying key technologies in Saskatchewan, Canada: evidence from patent information[J]. World patent information, 2011, 33(4): 364-370.

[5] YOON J, KIM K. Detecting signals of new technological opportunities using semantic patent analysis and outlier detection[J]. Scientometrics, 2012, 90(2): 445-461.

[6] ZHOU Y, LIN H, LIU Y, et al. A novel method to identify emerging technologies using a semi-supervised topic clustering model: a case of 3D printing industry[J]. Scientometrics, 2019,120(1): 1-19.

[7] LI M, PORTER A L, SUOMINEN A. Insights into relationships between disruptive technology/innovation and emerging technology: a bibliometric perspective[J]. Technological forecasting and social change, 2018, 129: 285-296.

[8] HE X, MENG X, DONG Y, et al. Demand identification model of potential technology based on SAO structure semantic analysis: the case of new energy and energy saving fields[J]. Technology in society, https://doi.org/10.1016/j.techsoc.2019.02.002

[9] PARK H, YOON J, KIM K. Identification and evaluation of corporations for merger and acquisition strategies using patent information and text mining[J]. Scientometrics, 2013, 97(3): 883-909.

[10] DING Y. Scientific collaboration and endorsement: network analysis of coauthorship and citation networks[J]. Journal of informetrics, 2011, 5(1): 187-203.

[11] CHOI S, KIM H, YOON J, et al. An SAO - based text - mining approach for technology roadmapping using patent information[J]. R&D management, 2013, 43(1): 52-74.

[12] YOON J, PARK H, SEO W, et al. Technology opportunity discovery (TOD) from existing technologies and products: a function-based TOD framework[J]. Technological forecasting and social change, 2015, 100: 153-167.

[13] 刘小玲,谭宗颖. 基于专利网络的技术演进研究方法探索[J]. 科学学研究, 2013, 31(5): 651-656.

[14] 胡正银,方曙,张娴,等. 个性化语义 TRIZ 构建研究[J]. 图书情报工作, 2015, 59(7): 123-131.

[15] 李欣,谢前前,黄鲁成,等. 基于 SAO 结构语义挖掘的新兴技术演化轨迹研究[J]. 科学学与科学技术管理, 2018, 39(1): 17-31.

[16] 陈亮,杨冠灿,张静,等. 面向技术演化分析的多主路径方法研究[J]. 图书情报工作, 2015,59(10): 124-130.

[17] 廖列法,勒孚刚. 基于 LDA 模型和分类号的专利技术演化研究[J]. 现代情报, 2017, 37(5): 15-20.

[18] YOON B, PARK Y. Development of new technology forecasting algorithm: hybrid approach for morphology analysis and conjoint analysis of patent information[J]. IEEE transactions on engineering management, 2007, 54(3): 588-599.

[19] YOON B, PHAAL R, PROBERT D. Morphology analysis for technology roadmapping: application of text mining[J]. R & D management, 2010, 38(1): 51-68.

[20] KIM Y G, SUH J H, SANG C P. Visualization of patent analysis for emerging technology[J]. Expert systems with applications, 2008, 34(3): 1804-1812.

[21] YOON J, KIM K. Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks[J]. Scientometrics, 2011, 88(1): 213-228.

[22] 汪雪锋,邱鹏君,付芸. 一种新型技术路线图构建研究——基于 SAO 结构信息[J]. 科学学研究, 2015(8): 1134-1140.

[23] HU Z, FANG S, WEI L, et al. An SAO-based approach to technology evolution analysis using patent information: case study-graphene sensors[J]. Journal of data and information science, 2015, 8(3): 62-75.

[24] 鲁若愚,银路. 企业技术管理[M]. 北京:高等教育出版社, 2006.

[25] 陈昌曙. 技术哲学引论[M]. 北京:科学出版社, 2012.

[26] 白光祖. 面向技术演化分析的语义化技术结构构建与应用研究[D]. 北京: 中国科学院文献情报中心, 2018.

[27] 王敏,银路. 技术演化的集成研究及新兴技术演化[J]. 科学学研究, 2008, 26(3): 466-471.

[28] 胡正银,方曙,文奕,等. 面向 TRIZ 的专利自动分类研究[J]. 现代图书情报技术, 2015(1): 66-74.

[29] ALTSHULLER G. 40 principles: TRIZ keys to innovation[M]. Norwood:Technical Innovation Center, Inc., 2002.

[30] 胡正银,方曙,隗玲. 基于SAO 的专利技术演化分析[C]//中国科学院成都文献情报中心. 贵阳:中国科学院成都文献情报中心,2015: 1-18.

[31] 付芸,汪雪锋,李佳,等. 基于SAO 结构的创新解决方案遴选研究——以空气净化技术为例[J]. 图书情报工作, 2019, 63(6): 75-84.

[32] KIM H B, HYEOK Y J, KIM K S. Semantic SAO network of patents for reusability of inventive knowledge[C]//2012 IEEE international conference on management of innovation & technology. Sanur Bali: IEEE, 2012: 510 - 515.
[33] CHOI S, PARK H, KANG D, et al. An SAO-based text mining approach to building a technology tree for technology planning[J]. Expert systems with applications, 2012, 39(13): 11443 - 11455.
[34] 翟东升,夏军,张杰,等. 基于专利特征抽取的技术树构建方法研究[J]. 情报学报, 2015, 34(7): 717 - 724.
[35] WANG X, QIU P, ZHU D, et al. Identification of technology development trends based on subject-action-object analysis: the case of dye-sensitized solar cells[J]. Technological forecasting & social change, 2015, 98: 24 - 46.
[36] 李倩. 基于专利的新兴技术弱信号识别方法研究[D]. 北京: 北京工业大学, 2014.
[37] 郭俊芳,汪雪锋,李乾瑞,等. 一种新型的技术形态识别方法——基于 SAO 语义挖掘方法[J]. 科学学研究, 2016, 34(1): 13 - 21.
[38] 王朝霞,邱清盈,冯培恩,等. 机械产品专利技术方案信息抽取方法[J]. 机械工程学报, 2009, 45(10): 198 - 206.
[39] 张艳. 汉语句法分析的理论、方法的研究及其应用[D]. 北京: 中国科学院自动化研究所: 2003.
[40] 王秀艳,崔雷. 采用混合方法抽取生物医学实体间语义关系[J]. 现代图书情报技术, 2013(3): 77 - 82.
[41] 张夫道,赵秉强,张骏,等. 纳米肥料研究进展与前景[J]. 植物营养与肥料学报, 2002, 8(2): 254 - 255.

作者贡献说明:
李晓曼:提出研究思路,设计研究方案,采集整理实验案例数据,撰写与修改论文;
张学福:确定研究思路、设计研究方案并进一步完善论文;
宋红燕:参与案例数据处理;
孙巍:进一步完善研究思路和方法流程,修改论文。

Research on Selection and Identification of Technical Elements for Patent Technological Evolution Analysis: A Case Study on Nano Fertilizer

Li Xiaoman Zhang Xuefu Song Hongyan Sun Wei
Agricultural Information Institute of Chinese Academy of Agricultural Sciences, Beijing 100081

Abstract: [Purpose/significance] Patent is one of the most reliable sources of technical intelligence. By patent analysis, one can realize the mining and the utilization of patent information, and the technology innovation. Technological evolution analysis refers to the process of emergence, development, transfer, change and even annihilation of technology themes. The focus of the current research is to deeply reveal the patent technology information, technical elements are the key to deeply reveal patent technology information. [Method/process] Proposing a method for identifying technical elements based on patent documents for specific fields by analyzing typical patents and feature recognition. [Result/conclusion] Applying our methods to the nano fertilizer field, through the analysis of typical patents, five technical elements are identified: materials, products, methods, functions and usage, and the identification of technical elements is completed based on SAO structure and domain vocabulary. Our methods can serve as a foundation for technological evolution analysis, and display field technical information from multiple perspectives.

Keywords: technological evolution technological elements SAO structure nano fertilizer technological network